

1

METHOD OF MAKING POWER MOSFET AND IGBT WITH OPTIMIZED ON- RESISTANCE AND BREAKDOWN VOLTAGE

FIELD OF THE INVENTION

The invention relates to the general field of power MOSFETs and IGBTs with particular reference to structures that optimize performance and methods for manufacturing same.

BACKGROUND OF THE INVENTION

For low power applications, metal-oxide-semiconductor field effect transistors (MOSFETs) normally follow a simple design with source, drain and gate arranged in the same plane. When such devices are produced in high power versions capable of handling hundreds of watts, different geometries are indicated. Such devices are generally referred to as power MOSFETs. Closely related to power MOSFETs are the IGBTs (insulated gate bipolar transistors) which are used in place of MOSFETs when large operational current is desired.

Since the geometries of the MOSFET and IGBT are very similar, we illustrate them here in a single figure (FIG. 1). Shown in schematic cross-section is an epitaxial silicon wafer made up of a high resistivity (typically about 500 ohm cm.) epitaxial N⁻ region 1 sitting atop a highly doped N⁺ or P⁺ silicon wafer 2. Sources 3 and 3a of the device are N⁺ regions which serve as sources for each pair of NMOSs and are directed into the plane of the figure. The inversion region of the NMOSs is shown as P type region 6 while the N⁻ epitaxial layer 1 serves as the extended drain region of the NMOSs for the power MOSFET or as the PNP transistor's base region. Gate oxide 4 extends above the neck region between two pairs of NMOS, overlapping slightly with region 6 on its two sides. Layer 5 is the gate electrode that rests on the gate oxide.

Thus, when sufficient positive voltage is applied to 5, an N type channel region is formed in 6 between 3a and 1 and between 3 and 1 and the device is turned on. In the case of a power MOSFET, current entering 1 (extended drain) flows down to region 2 which is an N⁺ silicon wafer upon which a metallic drain contact 7 is made. In the case of an IGBT, region 2 is P⁺ so the result is a combination of an nMOS device and a pnp transistor, with 7 being the emitter ohmic contact, 1 being the emitter, and 6 and its metal contact above being the collector, and region 1 being the base. Turn-on of the NMOSs supplies current to the base of the pnp transistor, thus turning it on.

While the design shown in FIG. 1 is widely used, it is not entirely satisfactory. The reason for this is illustrated in FIG. 2. In the case of a power MOSFET, when voltage is applied to the gate (layer 5), a depletion region is formed beneath it. The edge of the depletion region for an applied voltage of 100 volts is illustrated as dotted line 21 while the depletion edge for 200 volts is shown as dotted line 22. Once electrons have crossed from 3a into 1 by way of 6, they traverse the depletion region and enter the bulk of N⁻ region 1. In order for the breakdown voltage sustained between 5 and 1 to be kept as high as possible, the resistivity of 1 must be made as high as possible. Because of this, electrons traversing 1 encounter a series resistance which has been symbolized as resistor 26 in the figure. Series resistance of this kind is detrimental to device performance so a tradeoff has to be made between on-resistance and breakdown voltage.

Similar considerations apply to the IGBT design where a tradeoff has to be made between base resistivity and breakdown voltage. In practice, a fully satisfactory compromise

2

cannot be made so the result is an ineffective combination of nMOS and pnp BJT (bipolar junction transistor). The present invention presents a solution to this problem whereby both high breakdown voltage and low on-resistance can be achieved in the same device.

A search of the prior art revealed the existence of many references to devices of the type illustrated in FIG. 1. Some modifications to the doping profile of the N⁻ region are described with a view to optimizing the above discussed tradeoff, but they are severely limited as to what can be achieved. Typical examples are Lidow et al. (U.S. Pat. No. 5,742,087 April 1998) who form a relatively low resistivity epitaxial region on the surface prior to the formation of the source and gate, and Otsuki et al. (U.S. Pat. No. 5,714,774 February 1998) who use double emitters to obtain thyristor action as a way of reducing on-resistance.

SUMMARY OF THE INVENTION

It has been an object of the present invention to provide MOSFET and IGBT structures that have both high breakdown voltage and low on-resistance.

Another object of the invention has been to provide methods for manufacturing said devices.

These objects have been achieved by providing an N type shunt that extends from the N⁺ drain (for MOSFETs) or P⁺ emitter (for IGBTs), through the N⁻ region to a short distance below the gate oxide. To manufacture such a shunt, an epitaxial layer of N⁻ silicon is first provided with an N⁺ or P⁺ silicon substrate on its bottom surface. Through a suitable mask (contact or freestanding) on the top surface, the epi wafer is then subjected to bombardment by protons or deuterons. Because of ion transmutation doping, a region of N type material forms wherever the surface is not masked. By controlling the energies of the ions, this region can be caused to extend below the wafer's surface so as to just contact the N⁺ or P⁺ layer. Or the ions can simply be allowed to pass through the whole epi wafer. To convert shunt material in the immediate vicinity of the gate oxide back to N⁻ either counter-doping or trenching followed by dielectric filling and polishing are used.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic cross-section of a standard power MOSFET or IGBT.

FIG. 2 illustrates how the structure shown in FIG. 1 has an inherent series resistance built in.

FIG. 3 shows the structure taught by the present invention.

FIG. 4 shows how a region of N type material may be introduced through use of ion transmutation doping.

FIG. 5 is a plot of phosphorus concentration as a function of proton beam fluence.

FIG. 6 illustrates how a portion of the shunt may be converted to N⁻ material by means of counter-doping.

FIGS. 7 and 8 illustrate how the N⁻ conversion can be accomplished by etching a trench and then filling with dielectric material (normally oxide followed by polysilicon).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In order to short circuit the series resistance symbolized by resistor 26 in FIG. 2, a shunt 31 is introduced. As illustrated in FIG. 3, this is a region of N type silicon that extends from just inside layer 2 to just below the edge of the